Cloud Computing Benchmarking: *A Survey*

C.Vazquez, R. Krishnan, and E. John
Department of Electrical and Computer Engineering, The University of Texas at San Antonio,
San Antonio, Texas, U.S.A

Abstract – *Cloud computing gives service-oriented access to computing, storage and networking resource. Often, these resources are virtualized. The prospect of being able to scale computing resources to meet user demand has clearly caught the attention of developers and organizational IT leaders over the recent years. Considering the number of cloud computing providers and the different services each provider offers, cloud users need benchmark information that specifically addresses the unique properties of the cloud computing environment such as dynamic scaling. This paper compares five prominent tools (CloudCmp, CloudStone, HiBench, YCSB, and CloudSuite) that present workloads and/or methods for quantitatively comparing cloud computing offerings.*

Keywords: Cloud computing, Workload, Benchmarking, Performance evaluation

1 Introduction

The increase in popularity of cloud computing in recent years is driven by the advantages offered by the dynamically scalable, pay-as-you-go model. This enables organizations to focus on providing services to their customers while consuming the requisite computing resources as a utility. By eliminating the need for on-premises equipment, organizations avoid large capital expenses and instead focus resources towards faster deployment. The pay-as-you-go model allows an organization to grow naturally with customer demand. Since cloud computing resources scale elastically, utilizing cloud computing reduces the risk of over provisioning, wasting resources during non-peak hours, and reduce the risk of under provisioning, missing potential customers [32]. Success stories of start-ups like Instagram, which built-up a user base of over 150 million users in less than four years using only public cloud solutions [38], exemplify the potential for fast growth that utilizing cloud computing can provide.

Considering the number of cloud computing providers and the different services each provider offers, a customer shopping for an appropriate solution for their organization requires benchmark information that specifically addresses the unique properties of the cloud computing environment. A benchmark must provide an accurate representation of the workload the consumer intends on running. A benchmark targeting social networking sites should differ from a benchmark targeting database systems. Different applications running on the same computing platform can have different requirements in terms of computing, storage, and networking, and modern web applications can have wide disparities between peek and average demand [32]. A developer must ensure that the cloud provider's services can scale to meet their end-users' demand. Long response times from a cloud application can lead to limited adoption of an application since there are often competitors offering similar products.

Although standard methods for reporting the performance of cloud resources are still not available, tools have been suggested to give the consumer the ability to quantitatively compare the offerings of cloud providers. This paper identifies five such tools: CloudCmp [1], CloudStone [2], HiBench [3], YCSB [4], and CloudSuite [5].

2 Background

2.1 Cloud Computing

Cloud computing is a large-scale, distributed computing paradigm which is driven by economies of scale. Providers of cloud computing offer abstracted, virtualized, dynamically scalable, and managed resources on demand to external customers over the Internet [33]. These resources include compute, storage and networking. Cloud computing providers benefit from economies of scale in that they assemble massive datacenters operating tens of thousands of servers which service a wide customer base. Large-scale operation more effectively absorbs operational costs through the benefits of increasing the utilization of equipment, bulk discounts on purchased equipment, and reducing the cost of cooling and powering equipment [6]. The demand for large-scale computing resources continues to grow as Internet users generate larger sets of data to be processed.

The essential characteristics of cloud computing [7] are:

- On-demand self-service – The ability to provide computing capabilities as needed automatically, when needed.
- Broad networks access – Cloud services are available over the network and accessed through standard mechanisms.
- Resource pooling – Physical and virtual resources are dynamically assigned to serve multiple consumers using a multi-tenant model.
- Rapid elasticity – Capabilities are elastically provisioned and released quickly without perceived bound.
- Measured service – Cloud services automatically control resource use by leveraging appropriate metering capability (pay-per-use).

2.2 Virtualization

Virtualization is a fundamental component of cloud computing, allowing for pooling and dynamically allocating hardware resources. A server in a datacenter acting as a host machine is installed with a hypervisor which can simultaneously run instances of virtual machines or guest machines. These virtual machines are operating system instances managed by a separate controlling computer which loads them into respective host machines. With the controlling computer managing the computing resources of many servers, a cloud computing provider thus unifies the datacenter's resources into an encapsulated pool which can be allocated and released according to user demand.

2.3 Services

The NIST definition of cloud computing [7] categorizes the services that providers offer into three service models: infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), or a software-as-a-service (SaaS).

- An IaaS provides access to instances of unified resources including computing, storage, and networking. Providers offer flexible computing resources for a usage-based price. These resources are distributed as instances on demand which are treated like physical hardware. The user is left with the responsibility for demanding and initializing new instances when scaling is required.
- A PaaS provides many of the same resources as an IaaS but through an integrated environment which reduces the development burden of using the resources but also restricts features. PaaS providers offer a variety of computing and storage resources in a more constrained environment that can be accessed through APIs. Many application specific tools are pre-built and available to users such as web hosting, data management, business analytics, etc.
- SaaS, such as e-mail and Google Docs, are special-purpose software services which are used remotely by the end user. They are often built using PaaS and IaaS tools, but their implementation details are hidden from the end-user.

2.4 MapReduce

Since cloud computing now offers wide horizontal scaling, end-users are taking the opportunity to process massive sets of data, a service which was previously only available to users with a dedicated datacenter. Apache Hadoop [8], an open-source version of Google's MapReduce [9] and GFS [10], is a parallel processing framework used for many cloud-based batch-processing projects. A data set in a file system or a database is processed as follows:

1. Initialize - A list of key-value pairs is distributed over the nodes in a cloud.
2. Map phase – Each node performs a specified operation on the key-value pairs to produce new key-value pairs.
3. Shuffle phase – The new data is rearranged on the nodes according to a partition function which groups data.
4. Sort phase – Each node assigns new key-value pairs.
5. Reduce phase – Key-value pairs are merged to a data-set.

3 Cloud Benchmarking Tools

3.1 CloudCmp

CloudCmp is a proposed framework designed to estimate the performance and cost of a legacy application running on a cloud without the expense or effort of porting and deploying the application. To achieve this goal, CloudCmp uses an approach composed of three phases: service benchmarking, application workload collection, and performance prediction.

In the service benchmarking phase the services of six cloud providers (including Google AppEngine [11], Amazon AWS [12], Microsoft Azure [13], GoGrid [14], and Rackspace [15]) are selected based on their ability to provide cloud computing services necessary for web application development on a cloud. These cloud computing services include access to an elastic compute cluster, persistent storage, intra-cloud networking, and wide-area delivery networking. Each cloud service's performance and cost are estimated by running a collection of benchmarking tasks designed to exercise each of the characteristics of cloud computing services.

- Elastic compute cluster efficiency– Different compute clusters were tested with SPECjvm2008 [16] Java tasks. Java tasks were selected because of Java's portability. The performance of each cluster was measured by the finishing time of each task while the cost effectiveness was measured by the cost per task.
- Elastic compute cluster scaling – Scaling was measured by the latency between the time an instance was requested and when the instance was ready. The applicability of this metric is limited by the fact that not all services allow for scaling via instance request.
- Persistent storage services – To test the performance of a persistent storage service the latency to insert or fetch a random to and from a data table was measured. The test was carried out with table sizes of 1000 entries and 100,000 entries. The results showed that the operation and table size had a significant effect on the performance.

TABLE I
COMPARISON OF CLOUD BENCHMARKING TOOLS

	CloudCmp	**CloudStone**	**HiBench**	**YCSB**	**CloudSuite**
Target	Estimate the performance and costs of running a legacy application on a cloud	Capture "typical" Web 2.0 functionality in a cloud computing environment	Hadoop (MapReduce) programs including real-world applications	Performance comparisons of the new generation of cloud data serving systems	Characterize scale-out workloads
Cost	• Cost per task per instance type	• Cost per user per • month	• Not covered	• Not covered	• Not covered
Scaling	• Latency to allocate new instance	• Load balancer – Apache default or user defined	• None specific	• Scaleup • Elastic speedup	• None specific
Storage	• Latency to insert/fetch a random entry from pre-defined data table	• User's choice of relational database	• Aggregated bandwidth delivered by HDFS	• Adjust possible operations, data size, and distribution to target specific workloads	• Uses YCSB to assess serving systems
Networking	• Intra-cloud –TCP throughput between instances • Wide-area delivery network – send ping packets from distributed locations	• None specific	• None specific	• None specific	• None specific
Computing performance	• Latency of various SPECjvm2008 tasks	• Response time of request made by load generator	• Speed – job running time • Throughput – tasks completed per minute • System resources utilization	• Read/Update Latency	• Execution cycle profile • Instruction cache miss rate • IPC/MLP • Memory bandwidth utilization
Test environment	• Multiple instance types	• Amazon EC2 instances	• Hadoop cluster	• Data serving system	• Server
Service	• IaaS • PaaS	• IaaS	• PaaS	• PaaS	• IaaS
Workload	• User-defined application's request traces and each request's execution path	• Olio driven by Faban	• Sort • WordCount • TeraSort • Web search • Machine learning • File system	• Random operations on random data based on selected distributions	• Data serving • MapReduce • Media Streaming • SAT Solver • Web hosting • Web search

- Intra-cloud network – The available bandwidth between two instances in the cloud was tested by measuring the average TCP throughput of instances in the cloud using the iperf [17] tool for many pairs. This test is limited only to cloud providers which allow explicit intra-cloud communication.
- Wide-area delivery network – The latency of a cloud provider's delivery network was measured by sending ping packets from different geographic locations.

The goal of the application workload collection phase is to obtain a workload representation of a user's legacy application. It is proposed that this can be achieved by collecting the application's request traces and deriving an execution path for each request. In the performance prediction phase, the profiles of each cloud service and the workload representation of the legacy application would be used to estimate the total running time and total cost of running the application.

3.2 CloudStone

CloudStone is a toolkit for characterizing the workload of a typical social networking website. The goal of CloudStone is to give developers tools to investigate different implementation decisions which affect the performance and price of running a social networking website. These tools can currently only be utilized on a cloud service which can use Amazon EC2 instances. The three components of CloudStone are: Olio, automation tools for running Olio experiments, and a methodology for computing a suggested metric.

Olio features two complete implementations of a social-event calendar application and utilizes a time-varying workload generator, Faban [18]. The two application implementations, in both PHP and Ruby-on-Rails, provide an identical user experience allowing for a direct comparison of each development stack. Faban simulates multiple users simultaneously by running parallel agents on different which are controlled by one central coordinator. The central coordinator can also change the number of active users during a run. Faban also collects the latency of each request and utilization data.

Performing an experiment with CloudStone involves selecting a configuration for the Olio deployment, selecting a workload profile to be generated by Faban, and deploying the instances. The performance of the configuration of Olio will differ depending on the different tuning mechanism each implementation provides such as database caching, load balancer, etc. The results of the experiment are suggested to be expressed in terms of a metric of dollars per user per month.

3.3 HiBench

HiBench is a benchmark suite targeting the components of the Hadoop framework. The use of many realistic workloads fully exercises Hadoop's parallel computing component (MapReduce) and database component (HDFS). The benchmarking tasks selected can be categorized as micro-benchmarks, web search tasks, machine learning tasks, and HDFS benchmark.

- Micro-benchmarks include Sort [19], WordCount [20], and TeraSort [21]. Sort, which simply sorts a large collection of data, is intended to represent a class of MapReduce problem which transforms a data set. Similarly, WordCount is intended to represent a class which extracts a small amount of data from a large data-set. TeraSort is another sorting task but with a larger data-set. All of the micro benchmarks use tools included in HiBench to generate their input data-sets.
- Web search benchmarks, which include Nutch Indexing [22] and PageRank [23], test the ability to handle search-indexing systems. Nutch Indexing workload generates inverted index files from an input of web page links. PageRank calculates ranks of web pages according to the number reference links.
- Machine learning tasks include two workloads, Bayesian Classification and K-means clustering, from the Mahout library [24] which are used to test Hadoop's machine learning processing capabilities. Bayesian classification, a popular algorithm for data mining, is used on processed portions of Wikipedia [25]. The K-means algorithm, also popular for data-mining, is used to iteratively compute an approximation of the centroid of a multi-dimensional array which is randomly generated by HiBench.
- HDFS uses Extended DFSIO, an enhanced version of the DFSIO [26] program which is part of Hadoop. Extended DFSIO is file system benchmarks for finding the throughput of simultaneous read and write operations.

3.4 Yahoo! Cloud Serving Benchmark

Yahoo! Cloud Serving Benchmark (YCSB) is a tool developed by Yahoo! to benchmark their PNUTS [27] serving system. This benchmark focuses on scalable serving systems which provide read and write access to data. YCSB separates the task of benchmarking serving systems into two different tiers.

Tier 1 encompasses general performance as measured by the latency of a request when the database is under load. To test the balance of throughput and latency, the latency of a request is monitored as the throughput is increased. Tier 2 examines scaleup and elastic speedup, the serving system's ability to scale with increased load. This is achieved by observing the impact that adding more machines to the system has on the performance of the system. The ability of the system to scaleup well is described by system's latency remaining constant across multiple tests where the workload and server count are both increased. Elastic speedup measured test the impact of additional servers while a fixed size workload is running.

To test the performance and scalability of a serving system, YCSB uses a randomly generated workload instead of modelling a specific application. The YCSB client generates a dataset and operations according to a workload profile. The workload profiles contain user specifications for random distributions which are used to generate which operations will occur on which record.

3.5 CloudSuite

CloudSuite is a collection of benchmarking tasks which were used to characterize the inefficiencies in the micro-architecture of modern server CPUs used in a cloud computing environment. The benchmarking tasks were identified as some of the more common tasks which are handled using cloud computing. These tasks included data serving, MapReduce, media streaming, SAT solving, web hosting, and web search.

- Data serving – Cassandra [28] database exercised with a read-heavy YCSB workload.
- MapReduce - The Mahout library's Bayesian classification algorithm was run on a Hadoop cluster. The algorithm is used to process a portion of Wikipedia to guess the country tag for each article.
- Media streaming – The Darwin Streaming Server receiving request from simulated users generated by Faban.
- SAT solving – Cloud9 [30] parallel symbolic execution engine's Klee SAT solver
- Web hosting – CloudStone including Olio and Faban.
- Web search – Nutch/Lucene [31] index serving node receiving request from simulated users generated by Faban.

TABLE II
CLOUD COMPUTING BENCHMARK WORKLOADS

Target Application	**Workload**
Database	YCSB
Legacy application	CloudCmp
MapReduce	HiBench Mahout Bayesian classification
Media streaming	Darwin Streaming Service
Web 2.0	CloudStone

4 Conclusion

Cloud computing offers organizations the ability to scale to the size of their user base more efficiently and thus offers a competitive advantage if the proper services are selected. In this paper, we have presented available benchmarking tools for cloud computing services. CloudCmp offers an approach to benchmarking the individual cloud computing services offered by a provider. CloudStone provides a social networking application with simulated user interaction to test Web 2.0 applications. HiBench collects realistic workloads for the MapReduce processing framework. YCSB tests the performance and scalability serving systems with generated workloads. Finally, CloudSuite suggests workloads to capture the behaviour of the more common tasks in a cloud computing environment.

5 References

[1] Li, Ang, et al. "CloudCmp: comparing public cloud providers." *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, 2010.

[2] Sobel, Will, et al. "Cloudstone: Multi-platform, multi-language benchmark and measurement tools for web 2.0." *Proc. of CCA*. 2008.

[3] Huang, Shengsheng, et al. "The HiBench benchmark suite: Characterization of the MapReduce-based data analysis." *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on*. IEEE, 2010.

[4] Cooper, Brian F., et al. "Benchmarking cloud serving systems with YCSB."*Proceedings of the 1st ACM symposium on Cloud computing*. ACM, 2010.

[5] Ferdman, Michael, et al. "Clearing the clouds: a study of emerging scale-out workloads on modern hardware." *ACM SIGARCH Computer Architecture News*. Vol. 40. No. 1. ACM, 2012.

[6] Hennessy, John L., and David A. Patterson. *Computer architecture: a quantitative approach*. Elsevier, 2012.

[7] Mell, Peter, and Timothy Grance. "The NIST definition of cloud computing (draft)." *NIST special publication* 800.145 (2011): 7.

[8] Hadoop homepage. http://hadoop.apache.org/

[9] J. Dean and S. Ghemawat, “MapReduce: Simplified Data Processing on Large Clusters,” USENIX OSDI, December, 2004.

[10] Ghemawat, Sanjay, Howard Gobioff, and Shun-Tak Leung. "The Google file system." *ACM SIGOPS Operating Systems Review*. Vol. 37. No. 5. ACM, 2003.

[11] Google AppEngine. http://code.google.com/appengine.

[12] Amazon Web Service. http://aws.amazon.com.

[13] MicrosoftWindows Azure. http://www.microsoft.com/windowsazure.

[14] GoGrid Cloud Hosting. http://gogrid.com.

[15] Rackspace Cloud. http://www.rackspacecloud.com.

[16] SPEC Java Virtual Machine Benchmark 2008. http://www.spec.org/jvm2008

[17] Iperf. http://iperf.sourceforge.net.

[18] Faban Documentation. Sun Microsystems. 2009. http://faban.sunsource.net/0.9/docs/toc.html

[19] Sort program. Available in Hadoop source distribution: src/examples/org/apache/hadoop/examples/sort

[20] WordCount program. Available in Hadoop source distribution:src/examples/org/apache/hadoop/ examples/WordCount

[21] Hadoop TeraSort program. Available in Hadoop source distribution since 0.19 version: src/examples/org/apache/hadoop/examples/terasort

[22] Nutch homepage. http://lucene.apache.org/nutch/

[23] P. Castagna, “Having fun with PageRank and MapReduce,” Hadoop User Group UK talk. Available: http://static.last.fm/johan/huguk-20090414/paolo_castagna-pagerank.pdf

[24] Mahout homepage. http://lucene.apache.org/mahout/

[25] Wikipedia Dump. http://en.wikipedia.org/wiki/index.php?curid=68321

[26] DFSIO program. Available in Hadoop source distribution: src/test/org/apache/hadoop/fs/TestDFSIO

[27] B. F. Cooper et al. PNUTS: Yahoo!’s hosted data serving platform. In VLDB, 2008.

[28] The Apache Cassandra Project. http://cassandra.apache.org/.

[29] Darwin Streaming Server homepage. http://dss.macosforge.org/

[30] Liviu Ciortea, Cristian Zamfir, Stefan Bucur, Vitaly Chipounov, and George Candea. Cloud9: a software testing service. ACM SIGOPSOperating Systems Review, 43:5–10, January 2010

[31] Lucene homepage. http://lucene.apache.org

[32] Armbrust, Michael, et al. "A view of cloud computing." *Communications of the ACM* 53.4 (2010): 50-58.

[33] Foster, Ian, et al. "Cloud computing and grid computing 360-degree compared."*Grid Computing Environments Workshop, 2008. GCE'08*. Ieee, 2008.

[34] Grossman, Robert L. "The case for cloud computing." *IT professional* 11.2 (2009): 23-27.

[35] Binnig, Carsten, et al. "How is the weather tomorrow?: towards a benchmark for the

cloud." *Proceedings of the Second International Workshop on Testing Database Systems*. ACM, 2009.
[36] Alexandrov, A., et al. "Benchmarking in the Cloud: what it should, can, and cannot be." *4th TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC), VLDB*. 2012.
[37] Huppler, Karl. "The Art of Building a Good Benchmark." *Performance Evaluation and Benchmarking*. Springer Berlin Heidelberg, 2009. 18-30.
[38] Kavis, Michael J. *Architecting the Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS, and IaaS)*. John Wiley & Sons, 201